DOCUMENT RESUME

ED 293 884 TM 011 479

AUTHOR Keesling, J. Ward
TITLE Report on Analysis and Database Needs to Implement Model. Project: Multilevel Evaluation System.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Sep 86
GRANT OERI-G-86-0003
NOTE 20p.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Graphics; Data Analysis; Database Management Systems; *Databases; Educational Assessment; Elementary Secondary Education; Evaluation Methods; Microcomputers; Multivariate Analysis; *School Districts; *Students; *Teachers
IDENTIFIERS Multilevel Information System; *Univariate Analysis

ABSTRACT

This paper describes analytical concepts that can be applied to a comprehensive database containing information on students, teachers, and schools. The concepts are presented in terms of displays, potential inferences, and possible difficulties. Statistical techniques are mentioned, but not rigorously developed. Focus is on displays that can be produced by some microcomputer graphic programs. Each analysis is discussed in intuitive terms, relying on the common sense of the viewer of the display to guide his or her inferences. Topics covered include accumulating data, roster-based data presentations, desirable database system characteristics, analysis of single variables at one point in time, analysis of a single variable repeatedly measured over time, and analysis of the relationship between two variables. Two figures and seven graphs are presented. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED293884

Center for Student Testing, Evaluation,
and Standards

Project: Multilevel Evaluation System

Deliverable, September 1986

Report on Analysis and Database Needs to Implement Model

Project Director: Joan Herman

OERI-G-86-0003

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

LISA SIEGEL

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Center for the Study of Evaluation
UCLA Graduate School of Education
University of California, Los Angeles

TM011479

"The project presented, or reported herein, was performed pursuant to a grant from the Office of Educational Research and Improvement/Department of Education (OERI/ED). However, the opinions expressed herein do not necessarily reflect the position or policy of the OERI/ED and no official endorsement by the OERI/ED should be inferred."

ANALYTICAL CONCEPTS FOR AN LEA DATABASE MANAGEMENT SYSTEM

J. Ward Keesling

Introduction

This paper describes analytical concepts that can be applied to a comprehensive database containing information on students, teachers and schools. These concepts are presented in terms of displays, potential inferences, and possible difficulties. Statistical techniques are mentioned, but not developed rigorously. The intention is to discuss each analysis in intuitive terms, relying on the common sense of the person looking at the display to guide his or her inferences.

Accumulating Data in a Database

To enable any analyses to be performed, there must be a comprehensive database of information about students, teachers and schools. The basic data for students consist of records such as one might find in a "cumulative folder": test scores, courses taken, teachers for each grade level, special services received, comments from teachers or other school officials, attendance records, etc.

These data may come from a variety of sources within each district. The "Attendance Office" may be separate from the "Testing Office", for example, while other segments of the district bureacracy specialize in other data. It will be important to decide which of these offices will routinely supply data, and for them to agree to a common system of identifying students to be used throughout the district.

The students' actual names are usually an inconvenient identifier: many characters have to be set aside in each record for the name; names are not necessarily unique; students may change their names; it is easy to make key-entry errors for names; and using some other identifier may provide a bit more confidentiality for the data. A numerical identifier for each child should be developed and a separate file linking this identifier with the child's name (or history of names) should be maintained.

The common identifier makes it possible to merge the files of data obtained from the different offices over time. Cooley (1986) points out that mismatch rates as high as 20% result when careful attention to identifying students is lacking.

Roster-Based Data Presentations

A basic presentation of this data could be thought of as a student roster in which a selection of the data in the database is listed next to each student's name. This can be a very useful presentation for the teacher, who could, for example, use the multiple indicators in the report at the beginning of the year to group students for instruction. It can be helpful to both the teachers and the staff maintaining the database as a means to spot errors. Teachers may recognize particular data values as highly unlikely (e.g., an especially low score for a high-achieving student) and could flag them for examination as possible errors. It is wise to include a review of this type in the development of the database.

Once this information is in the computer, then most of the analyses to be discussed can be performed. For some analyses, aggregates of the

student data such as classroom average score or school average score will be calculated first. To complete these analyses, information about teachers and schools will also be needed.

Information about teachers could include the amount of homework they assign, their courses beyond the Bachelor's Degree, their years of teaching experience, etc. Information about schools could include the ambient noise level, the age of the building, the lighting, the size of classrooms, and the number of students per class, etc.

Any information that could be presented in the form of a roster (as described above: a list of values attached to the identification of an individual or place) can be used in the analyses to be described in this paper. To do any of these analyses, the data must first be in the computerized records.

Desirable Database System Characteristics

The database system that contains and manages all of this information should provide several functions that will make it easy and flexible to use. It should facilitate the creation of attractive data entry and report formats. Cooley (1986) indicates that he has reproduced even the color of the standard forms used in a district, so there is very little new learning for the personnel entering the data.

It is also important that the database system have the capability to subset the records by school building, grade level, teacher, special program, etc. If information about a student is stored in several different files, then the database system must be able to create "views" of the data that merge information from these files using the common identification scheme referred to earlier.

Most database management systems do not provide an integrated statistical analysis system. Almost all do provide the mean and standard deviation for each subgroup called out in a selection of data. In many cases this will be enough information to create interesting and interpretable displays. However, sophisticated hypothesis testing will require additional software.

We focus our discussion on displays that can be produced by some microcomputer graphic packages. Note, however, that most database management systems do not include a graphics package. Integrating available graphics packages with a database system will require additional software. In addition, some of the displays to be discussed will need to be programmed. Useful code for some of them can be found in Velleman and Hoaglin (1981).

Modes of Analysis

Each of the displays to be described can be used in several ways. Three of the most common are:

Description: To show what the pattern of a particular set of values or a relationship looks like. The description can set a context for a discussion of goals, policy and activities. Showing a large proportion of scores below the first quartile may help to motivate remedial activities. Showing that homework is related to achievement may help to direct policy.

Hypothesis Generation: To highlight features of the data that might deserve further examination. Displays could point out a problem (Why are the 6th grade scores at PS-112 so low?), or highlight a success (What did Mrs. Smith do that her class scored so well?).

Confirmation: To test hypotheses statistically to demonstrate that a relationship exists or that a specific program is effective.

In the discussion of each type of analysis, examples will be provided of these three modes.

Three Types of Analysis

Three types of analysis to be discussed are:

Analysis of a single variable at one point in time: e.g., test scores of sixth graders in 1986.

Analysis of a single variable measured repeatedly in time: e.g., test scores of sixth graders from 1976 to 1986.

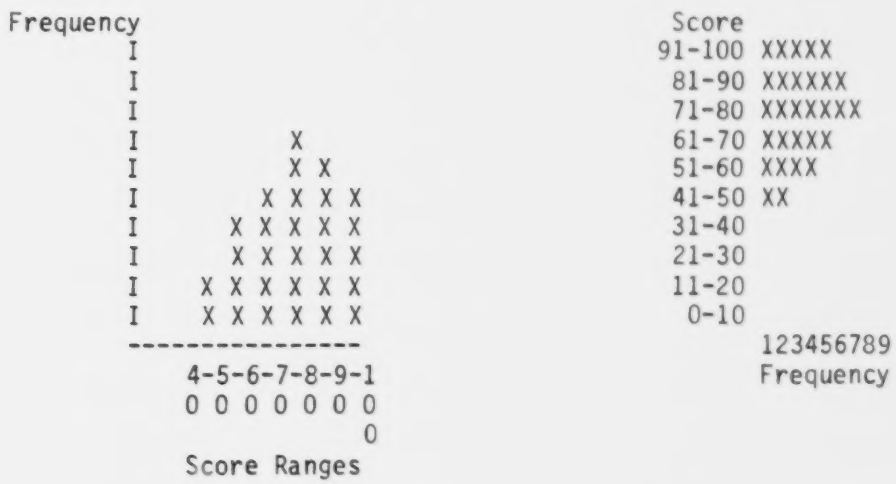
Analysis of the relationship between two variables: e.g., the relationship between amount of homework completed and test scores.

The discussion of each type of analysis will present displays of information that might be used to describe the data. This will be followed by a discussion of the aspects of these displays that could be examined and statistical summaries that capture this information. Next, enhancements of the displays that can be used to show the status of several groups simultaneously will be presented, and statistical techniques for conducting confirmatory analyses will be outlined.

Univariate Analysis at a Single Point in Time

Typical displays for a single group data (such as test scores of 6th graders in 1986):

Frequency Histograms (vertical or horizontal):



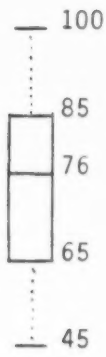
Stem and Leaf Plots:

Scores

Stems (10's)	Leaves (1's)
10:	0
9:	2789
8:	233568
7:	1246779
6:	25678
4:	56
3:	
2:	
1:	

(The data value is the sum of a leaf plus ten times the stem.)

Box Plots:



Some of the features we might look for include:

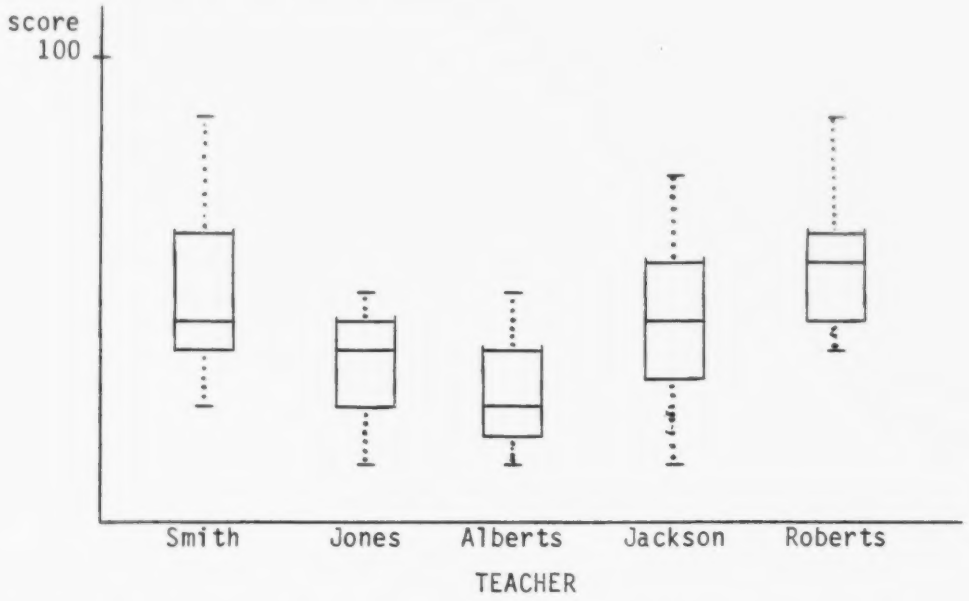
The general location of the scores. Is the typical value of these scores the sort of value we expect or desire? The statistical summary appropriate here would be the mean or median value. For the data shown, the mean is 74.3 and the median is 76. Note that the stem-and-leaf plot preserves all of the original data, so that one can compute these values exactly only from this display.

The spread of the scores. Do these scores represent a lot of variation in student performance, or only a little? Statistical summaries appropriate here are the interquartile range and the standard deviation (or the variance, which is the square of this value). Again, the stem-and-leaf plot has preserved the exact raw data, so these values may be computed precisely only from this display.

Asymmetry. Is there a large shift to one end of the scale or the other? A shift to the top end of the scale might result from a class being run on principles of mastery learning.

The box plot is constructed from five summary numbers: the maximum score, the minimum score, the first and third quartile scores and the median. The box shows where the middle 50% of the distribution lies. This type of display can summarize very economically the distribution of

a group of scores. It is probably the most adaptable of these displays to showing multiple groups of scores simultaneously. Imagine a set of boxes placed on the same vertical scale, each showing the results of a different classroom of 6th graders:



This display is very effective at showing how the classrooms did compared to one another. It is much easier to interpret than a set of histograms or a set of stem-and-leaf plots, although it does lose some of the information in the latter display. It should be clear that many box plots can be put into the space that only a few of the other displays would fill, enabling one to describe many more groups in one easy-to-read display.

Some computer packages permit side-by-side histobars of different colors or shadings, but these prove to be difficult to interpret if there are more than two groups.

A t-test may be used to confirm differences between two groups, or an analysis of variance (F-test) may be used as an omnibus test of

differences if there are more than two groups. Alternatively, when there are more than two groups, one can use multiple comparison procedures to make inferences about which groups are different from which others.

Examples of confirmatory analyses include: contrasting male and female math scores to confirm that female students are not achieving as well as the males; or contrasting reading scores of Chapter 1 and non-Chapter 1 students to confirm that the Chapter 1 students are in need of remediation. The hypotheses to be confirmed should be based on theory or previous data, not the displayed data -- to do otherwise is akin to peeking at the tossed coin prior to placing a bet.

As useful as analyses of scores at a single time point may be, trends over time are more revealing of the output of the school system. We turn next to displays and analyses suitable for such data.

Longitudinal Analyses

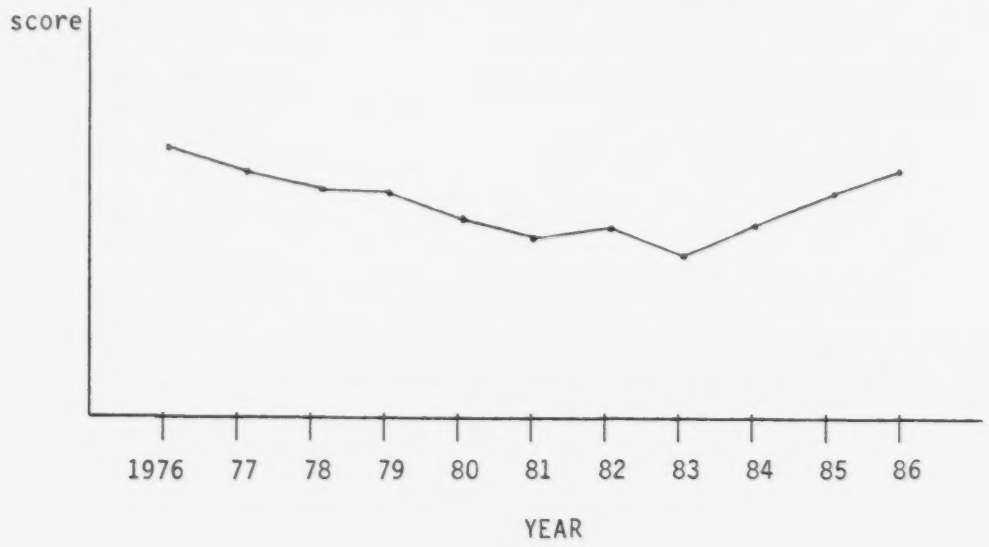
Longitudinal analyses may be of two types: unit repetitive and unit replicative. The type of analysis will not influence the display, or (in most cases) the statistical analysis, but it will influence the sense we make of the information.

Unit repetitive analyses measure or assess the same physical entities through time. For example, tracing the test scores of the same cohort of students from the time they are in first grade to their graduation from High School.

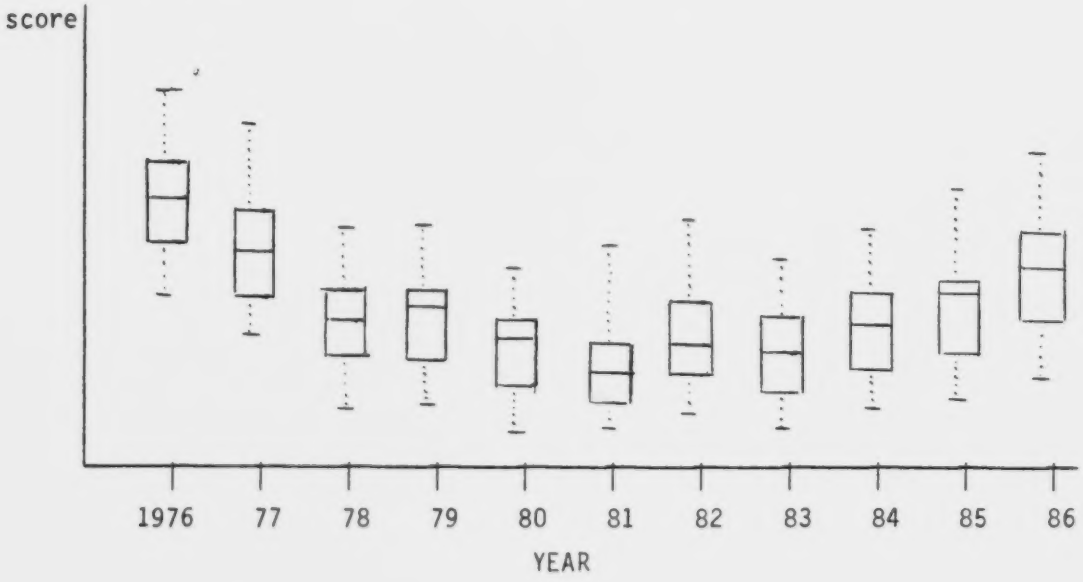
Unit replicative analyses trace the same conceptual entities through time. For example, the scores of sixth graders in the district from 1976 to the present. The performance of sixth graders is measured each year -- on a different group of students. So this is a conceptual entity, not a physical entity.

A more difficult example would be the scores of Mrs. Smith's sixth grade class from 1976 to 1986. Mrs. Smith is the same person and may very well teach in the same room for all those years. So part of the "unit" repeats, while part (the students) replicates. If we wish to make inferences about Mrs. Smith, we have to recognize that the trace of the performance of her sixth grade classes may reflect variation in demographic or other effects carried by the students, not having to do with Mrs. Smith. In contrast, when we examine the performance of a cohort of student over time, we are looking at the record of the cumulative influences on them.

The typical display for this type of analysis is a chart where the horizontal axis is time, and the vertical axis is the value of interest. A mark is placed where each time point intersects the obtained value. Sometimes adjacent points are connected by lines to emphasize the trace through time. (Of course, it should not be assumed that the values in periods between measures fall on the connecting line.)



Another display would be to replace the mark with a vertically oriented box plot at each time point. This would be especially valuable if there was more than one unit being followed through time (e.g., the box plots could be for scores of all sixth graders at each point in time). This display provides more information about the distributions of scores over time.



In these displays we would be looking to see if there is a slope to the line showing growth or decline (e.g., higher test scores, declining absentee rates). Our common sense would indicate whether this is a favorable trend, or whether it is moving slowly (shallow slope) or rapidly (steep slope).

If there is no trend (no slope), then we have to assess whether the average value is indicative of what we would want or expect.

There are several pitfalls in interpreting these trend charts. In charting trends in test scores, for example, one has to be careful of changing metrics. Changing the test publisher, or even changing the form

or level of a test can change the metric. The change in metric might masquerade as a change in performance, or mask a true change.

Another problem is in giving too much weight to the last change. It is too easy to see the light at the end of the tunnel of declining test scores when they went up one point at the last testing. This does not necessarily establish a new trend!

Typical statistical summaries involve the computation of the slope of a straight line "fitted" to the points, and its intercept. Sometimes the trend is better explained by a curvilinear relationship to time and a more complex function has to be fitted.

If there has been an intervention of some kind (such as the start of a compensatory education program), then trend lines might be computed prior to and after the intervention to determine if the slope and/or intercept have been affected. These comparisons become rather complex, and have to take account of the correlations among the values from one time to the next.

Trend lines for several groups may be displayed simultaneously on the same chart, although this can quickly become chaotic. Even using different forms of line (dashes, circles, etc.) does not help if there are many groups.

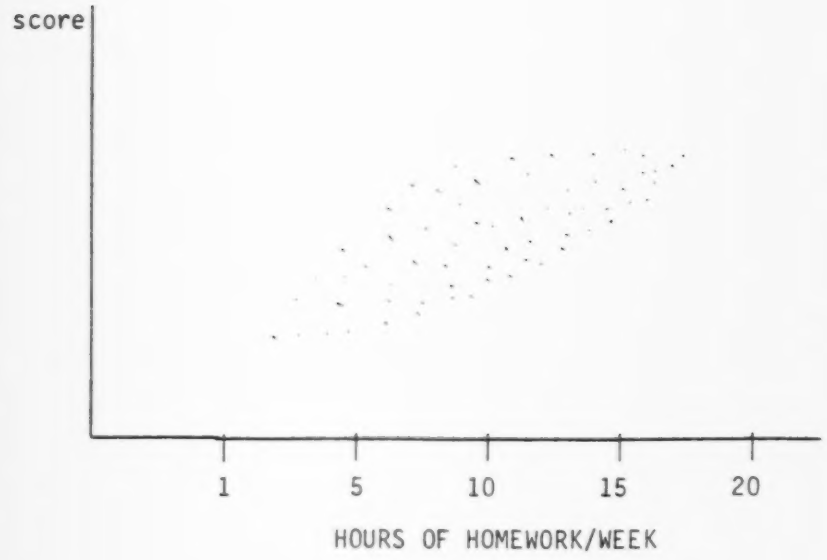
Statistical comparisons of two or more groups over time usually involve comparing the shapes of the trend lines. If the shapes are alike (e.g., parallel lines), then the intercepts are compared to see if there is a constant difference among groups over time. For unit replicative designs, a standard group-by-time analysis of variance will yield the desired statistical tests: the group-by-time interaction is a test of

parallelness of shapes, the group contrasts (in the absence of interaction) are combined for a test of constant differences, and the time contrasts (also in the absense of interaction) can be used to assess the degree of curvature in the trends. If interaction is detected, then each group must be described by its own trend line and comparisons among groups are not consistent over time.

For unit repetitive designs, the repeated measures on each student unit must be transformed into "trend scores" that capture the nature of that unit's trend line. These trend scores are used to compare groups (using the pooled, within-groups variance as the error term). If any trend score other than the one for constant (i.e., no trend) shows statistically significant difference across groups, that is evidence of interactive effects. If the constant score is the only one that is statistically significant, then the groups are consistently different over time. If all the group contrasts are non-significant, then the grand mean on the trend scores may be examined to establish the nature of the over-all trend.

Relationships Between Two Variables

A scatterplot of data points is the most effective way to display the relationship between two variables. For example, consider the relationship between the hours of homework a student does each week and his/her score on the Spring achievement test. This relationship can be shown by plotting the scores versus the hours of homework.

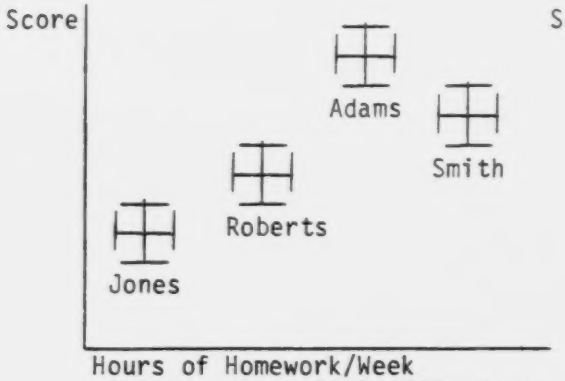


Typical numerical summaries of this information consist of the correlation coefficient (which measures the degree of relationship and tells whether it is positive or negative), the regression coefficient (a measure of the slope of the best-fitting straight line) and the intercept (showing where the best fitting straight line crosses the vertical axis).

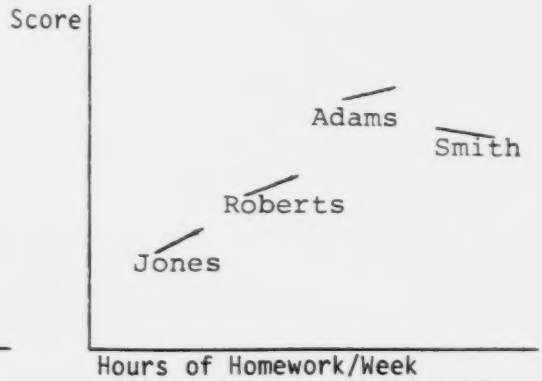
These numerical summaries may be treated as simply descriptive, or may be used in confirmatory analyses to show that a relationship expected by theory is (or is not) confirmed by the data.

When there are multiple groups of students involved, the scatterplot can take the form of locating "potent" crosses at the point on the plot representing the group means on the two variables. The arms of the crosses can represent a measure of the spread of scores on the two variables -- for example, from the first to the third quartiles (the same spread marked off by the box part of the boxplot).

Another type of display for groups would be to show a segment of the regression line for each group corresponding to the span between that group's first and third quartile scores on the predictor variable (the horizontal axis).



Crosses show middle 50% of the distribution of homework hours and score for the students in each teacher's classroom.



Line segments show direction of the relationship within each classroom. The location of the segment along the horizontal axis shows the middle 50% of the hours of homework reported by the students in that class

Analyses of these displays can be quite complex. For example, it is useful to know if the relationship is the same in all groups. If it is not, then a simple description of the typical relationship found within the groups will not be accurate.

Some analysts examine the effects of grouping by contrasting the typical within-group regression line to that formed by considering the group means. For example, it may be that hours of truancy counselling improves individual attendance within schools, but the total hours of such counselling at a school may be negatively related to average school attendance. Are counselling resources being directed to the schools that need it most, but proving insufficient to make large changes in attendance at the neediest schools? Or, are disruptive students being counselled to return to school, causing others to stay home?

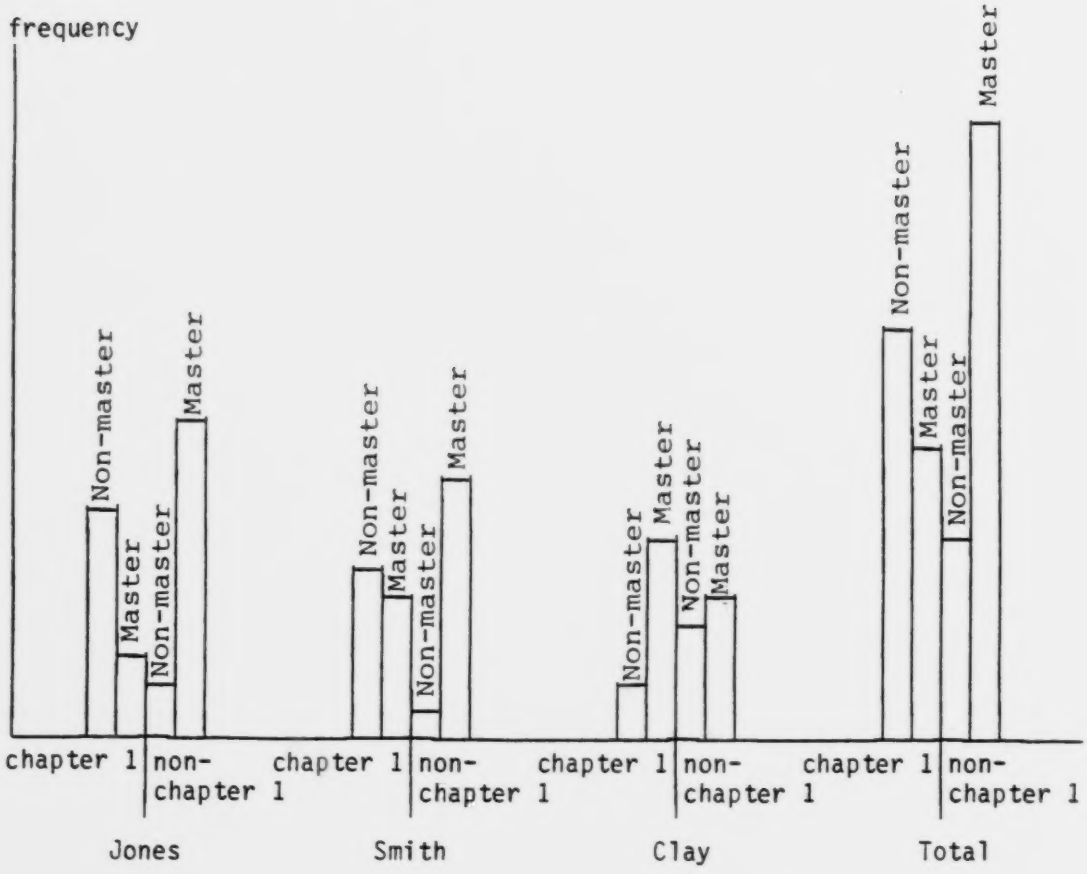
Relational analyses are very difficult to interpret at face value. Many explanations may arise to suit any particular display of data. The decision maker will have to understand all aspects of the information presented and be willing, in some cases, to ask for additional information before rushing to act.

When the variables being related to one another are not amounts, but classifications (e.g., male or female, black or white, Chapter 1 or non-Chapter 1, master or non-master of an objective) then crosstabulations of counts of students may be used to show relationships. For simple cases, raw tabulations of numbers may be fairly easy to interpret (if there are two categories in each variable, say).

Generally, we are trying to determine whether classification on one variable is dependent upon the classification on the other variable. For example, we might want to know if a larger or smaller percentage of Chapter 1 students has mastered basic addition facts, compared to non-Chapter 1 students.

When several groups must be compared (e.g., tabulating master/non-master versus Chapter 1/non-Chapter 1 in each school) interpretation becomes more difficult.

Simple plots of histobars showing frequencies in the cells of classification may be difficult to interpret when there are very many of them. Confirmatory analyses for crosstabulated data may be performed by using chi-square tests of relationships, or by using the more recently developed techniques of log-linear models. Discussion of these methods is beyond the score of this paper.


frequency
Non-master
Master
Non-master
Master
Non-master
Master
Non-master
Master
Non-master
Master
Non-master
Master
Non-master
Master
chapter 1
non-chapter 1
chapter 1
non-chapter 1
chapter 1
non-chapter 1
chapter 1
non-chapter 1
Jones
Smith
Clay
Total

REFERENCES

Cooley, W. & Bickel, W. (1986). Decision Oriented Educational Research, Kluwer Nijhoff Publishing, Boston.

Velleman, P., & Hoaglin D. (1981). Applications, Basics, and Computing of Exploratory Data Analysis. Duxbury Press, Boston.